

Self-organized manifold learning and heuristic charting via adaptive metrics

D. Horvath

*Centre of Interdisciplinary Biosciences, Faculty of Science,
Jesenna 5, P. J. Safarik University, 04154 Kosice, Slovak Republic*

J. Uličný, B. Brutovsky

*Department of Biophysics, Faculty of Science, Jesenna 5,
P. J. Safarik University, 04154 Kosice, Slovak Republic*

Classical metric and non-metric multidimensional scaling (MDS) variants are widely known manifold learning (ML) methods which enable construction of low dimensional representation (projections) of high dimensional data inputs. However, their use is crucially limited to the cases when data are inherently reducible to low dimensionality. In general, drawbacks and limitations of these, as well as pure, MDS variants become more apparent when the exploration (learning) is exposed to the structured data of high intrinsic dimension. As we demonstrate on artificial and real-world datasets, the over-determination problem can be solved by means of the hybrid and multi-component discrete-continuous multi-modal optimization heuristics. Its remarkable feature is, that projections onto 2D are constructed simultaneously with the data categorization (classification) compensating in part for the loss of original input information. We observed, that the optimization module integrated with ML modeling, metric learning and categorization leads to a nontrivial mechanism resulting in generation of patterns of categorical variables which can be interpreted as a heuristic charting. The method provides visual information in the form of non-convex clusters or separated regions. Furthermore, the ability to categorize the surfaces into back and front parts of the analyzed 3D data objects have been attained through self-organized structuring without supervising.

Manifold learning (ML) [1] is a technical term for a group of techniques developed to reduce dimensionality of high-dimensional data, facilitating their eventual visualization, evaluating and understanding at intuitive level. The ML approach usually serves as a preprocessing step in familiarizing with data and for the formulation of hypotheses leading to further data analysis. It has found many important applications in biology, robotics or visual data mining. In recent years the ML techniques have also found applications in the physics of deterministic chaos [2], as well as techniques for extracting structural information from X-ray diffraction snapshots [3].

Projection from a higher to lower dimension is not straightforward and must meet the ultimate requirement of capturing the essence or patterns of information content of high dimensional datasets. Hence, the ML methods differ markedly in the kind of the original information they are required to preserve during the transformation. From this point of view, the basic classification of ML methods into local and global should be mentioned.

The representative of the global methods are well known *principal component analysis* (PCA) [4, 5] or classical MDS [6, 7] *non-linear multidimensional scaling* (MDS) [8–10]. The latter approach, MDS, is studied and extended in the present work. Both the MDS variants, non-metric as well as metric, are formulated as stress minimization problems where the stress is defined in the terms of differences between pairwise dissimilarities of data points and distances of assigned projected coordinates [11]. Many methods have been developed that incorporate the conservation of

the quantities such as distances and generalized distances. In situations where the topological concepts offer more feasible projections, pure Euclidean description is abandoned in favor of more flexible geometries. For example, the method Isomap [12, 13] uses geodesic instead of Euclidean distances and can be thus preferentially applied to nonlinear manifolds.

Elementary ML methods face serious difficulties when confronted with noisy [14] or intrinsically irreducible high-dimensional datasets. In such cases, projection composed of approximate local isometrics is usually constructed. In many real-world applications, one often tackles general manifolds, where the domain decomposition and segmentation problems often occur. These are typical for the closed manifolds, intersecting circles, sphere surfaces or non-orientable surfaces such as Möbius strip or Klein bagel, which are analyzed later in this paper. The ML segmentation tasks producing non-overlapping domain decomposition are also known as identifications of *charts* and *atlases* [15]. The decomposition is NP hard local categorical assignment, which occurs in the graph coloring or graph partitioning problems.

Until now, many alternative interdisciplinary approaches based on different principles have been developed to deal with the ML data preprocessing. The famous *stress function concept* has been first introduced by Kruskal [6, 7]. Consequently, the scientists [16] are turning their attention to new ideas and methods derived from the connection of ML with mathematical and physical modeling or nature-inspired optimization sciences [17]. As an example may serve physically inspired ML method [18]. Another method called *elastic map* exploits the mechanical analogy with the system of the elastic springs. The concept called *diffusion map* [19, 20] introduces diffusion distance less susceptible to the noise compared to Euclidean measures. The approach *relational perspective map* [21] consists in using parallels between mutual relations of data items and the behavior of the positions of the charged particles, which are repelling each other but are simultaneously confined to toroidal surface. Several new approaches to perform ML [22] including kernel regression [17] take inspiration from the evolutionary general-purpose heuristics and genetic algorithms. In [11] the class of the models based on the generalized B-C energy (stress) functions [23, 24] which has its origins in the optimization approach of Sammon [16].

Rapidly increasing computer power allows to tackle ML problems in previously unexpected ways. Further progress in the ML and MDS techniques can be made by applying simple, but computationally demanding *optimization-based approaches*. In this respect, we propose the variant of data adaptive metrics aimed to provide combined description in the terms of continuous/quantitative and categorical variables. Simultaneous use of discrete (categorical) and continuous variables needs heuristics-based optimization to avoid getting stuck in the local optima. As demonstrated in the below presented simulation examples, the advanced hybrid construction enables to adapt the distance metrics simultaneously with the categorical classification and adjustment of the continuous projected coordinates.

The ML process may be also viewed as a kind of stochastic optimization which is inspired by the imitation of natural systems. Designers of optimization techniques usually imply biologically-inspired concepts [25] to discover suitable rules [26]. On the other hand, many heuristics-based optimization methods benefit from the analogies between the optimization dynamics and physical processes as well [27].

In the paper we present computational results of heuristic simulation MDS technique. We assume, that the combination of a few existing optimization approaches may provide better results than the only method. The stochastic optimization method we applied combines the advantages of *grid search* (GS) [28], *extremal optimization* (EO) [29], and *hysteretic optimization* (HO) [30]. As demonstrated below, dynamical behavior of the above combination produces

very interesting behavior. In the next we give a brief description of the respective methods and their benefits.

GS is the standard way of performing exhaustive optimization in the hyperparameter space. But this strategy does not scale well for large problems. GS is an efficient in a one-dimensional or two-dimensional domains since the problems at higher dimensions occur due to "the curse of dimensionality".

The problem of getting stuck in local optimum is partially solved by incorporating EO method, inspired by the stylized model of coevolutionary process proposed by Bak and Sneppen [31]. Since then, many authors have extended the method (see e. g. [32]) and applied it in many contexts [33]. The essence of the method consists in the identification of the low-quality components and their subsequent elimination. The method exploits highly nonlinear mechanism of the large fluctuations - avalanches, known to be efficient in the exploration of many local optima and overcoming of the barriers in the search space [34]. The main difficulty with the EO applications is, that its implementation necessitates specific definition of the local fitness (scoring, objective) function.

HO method [35] is inspired by the mechanism of the global reordering during the demagnetization of magnetic samples due to damped alternating magnetic field. Nevertheless, the method can be formulated more abstractly and adapted to non-magnetic problems as well. The HO method provided successful outcomes in the case of the benchmark *traveling salesman problem* [36]. We justify the below presented ML application of the HO type technique by the fact, that suggested type of distance metrics involves global parameter with system-wide impact which can be roughly regarded as analogous to the intensity of external magnetic field.

The paper is organized as follows. In the section I we describe the formulation of MDS with the use of the adaptive metrics. In section II we discuss the optimization strategies appropriate for given purpose. The datasets and corresponding numerical results illustrating our approach are described in sec.III. Finally, the conclusions are presented.

I. MDS WITH ADAPTIVE METRICS

Below we analyze N data items, each having D ($D > 2$) components (column features, or classes)

$$\{ \mathbf{X}_i \in \mathbb{R}^D, i = 1, 2, \dots, N \}, \quad (1)$$

where $\mathbf{X}_i = [X_{i,1}, X_{i,2}, \dots, X_{i,D}]$. Regarding MDS technique, essential information is comprised in $N \times N$ elements of dissimilarity matrix $d_{i,j}^{(D)}$. In here presented specific application, $d_{i,j}^{(D)}$ obtains standard Euclidean form $d_{i,j}^{E,(D)} = \sqrt{(1/D) \sum_{z=1}^D (X_{i,z} - X_{j,z})^2}$, but other choices are possible as well.

The process of dimensional reduction onto dimension $P < D$ can be viewed as ongoing iteration of the configuration tuples $C(t) \in \mathbb{R}^P \times \Omega \times \mathbb{R}^N$ including N data points

$$\begin{aligned} C(t) &\equiv \{ [\mathbf{x}_1(t), s_1(t)], [\mathbf{x}_2(t), s_2(t)], \dots, [\mathbf{x}_N(t), s_N(t)] \}, \\ \mathbf{x}_i(t) &\in \mathbb{R}^P, \\ s_i(t) &\in \Omega \equiv \{ 0, 1, \dots, N_s - 1 \} \end{aligned} \quad (2)$$

where $i = 1, 2, \dots, N$; the discrete time t ranges from 0 to t_* ; $C(t)$ consists of the system of N vectors of P real valued Cartesian coordinates $x_{i,1}(t), \dots, x_{i,P}(t)$. One of the cornerstones of the proposed approach is, that the uncertainty

and frustration which arose from a projection effect may be reduced by introducing categorical variables $s_i(t)$, N_s being the number of their possible values.

At the heart of the MDS ML approach stands the requirement of approximate fulfillment of $N(N-1)/2$ conditions after the stop time t_*

$$d_{i,j}^{(P)}(t^*) \equiv d^{(P)}(\mathbf{x}_i(t_*), s_i(t_*), \mathbf{x}_j(t_*), s_j(t_*)) \simeq d_{i,j}^{(D)}, \quad (3)$$

which approximate $d_{i,j}^{(P)} = d_{j,i}^{(D)}$. Since the conditions of distance-preservation are too demanding to be achieved in all the eventual applications, it is desirable to solve an approximation problem by iterative optimization. The quality of the approximation may be assessed through the absolute error term

$$e_{i,j}(t) \equiv \left| \frac{d_{i,j}^{(P)}(t) - d_{i,j}^{(D)}}{d_{i,j}^{(D)}} \right|$$

but more appropriate case-dependent scoring variants and weighting schemes may be devised for specific situations. The effort is to achieve trajectories revolving around desired outcome $e(\mathbf{x}_i(t_*), s_i(t_*), \mathbf{x}_j(t_*), s_j(t_*)) = 0$.

In analogy with the well known additive interaction effects we assume, that required properties of the i -th projected component may be attained by checking the values of the local potentials constructed as

$$V_i(C) = \frac{1}{N-1} \sum_{j=1; j \neq i}^N e_{ij}. \quad (4)$$

The overall views about the system performance and convergence can be obtained by minimizing the total potential

$$V_{\text{tot}}(C) = \frac{1}{N} \sum_{i=1}^N V_i(C). \quad (5)$$

Note, that the term stress function is more commonly used within the ML MDS context [11, 16]. When seen from the point of view of Bak-Sneppen model [31], the value V_i plays role of the fitness. Being inspired by Monte Carlo simulations of the spin systems, the overall categorization dynamics was characterized by calculating the instant "magnetization"

$$\text{Mag}(t) = \frac{1}{N} \sum_{i=1}^N s_i(t). \quad (6)$$

In the presented version of MDS algorithm we propose parametric distance measure

$$d^{(P)}(\mathbf{x}_i, s_i, \mathbf{x}_j, s_j) = (1 + H|s_i - s_j|) d^{\text{E},(P)}(\mathbf{x}_i, \mathbf{x}_j). \quad (7)$$

Instead of relying on pure Euclidean distance $d_{i,j}^{\text{E},(P)}$, we use modification with the multiplicative factor $1 + H|s_i - s_j|$ that is supposed to improve the matching according Eq.(3). Here, $H \in \langle H_D, H_U \rangle \subset \mathbb{R}$ is the real valued parameter. Its global system impact motivates the use of the HO optimization. The dependence on $|s_i - s_j|$ represents the interaction due to differences in categories. When the different data items are differently categorized ($s_{i \neq j} \neq s_j$), the Euclidean distance $d_{i,j}^{\text{E},(P)}$ changes in the positive or negative sense according to the sign of selected H parameter. The matching of the categories ($s_i = s_j$) simply yields basic choice $d_{i,j}^{(P)} = d_{i,j}^{\text{E},(P)}$. Since not only $\{\mathbf{x}_i\}_{i=1}^N$, but $\{s_i\}_{i=1}^N$ and H are unknown as well, the approach constitutes complex inverse problem which requires simultaneous tuning of distance, category and metrics. This optimization problem is solved by combining beneficial features of the three mentioned

optimization methods: GS, HO and EO. As demonstrated in our numerical experiments, the iterative procedure incorporating them can exhibit very complex dynamics and behaviors. The optimization methods are considered to have access to different subsystems: (i) GS method is applied to optimize $\{\mathbf{x}_i\}_{i=1}^N$ and $\{s_i\}_{i=1}^N$; (ii) the parameter H is optimized by the *self-organizing dynamics* based on the modified HO method; (iii) EO applied to vary $\{s_i\}_{i=1}^N$ to disentangle partially improperly justified categories. The self-organization arises through decentralized interactions without primary knowledge of the way how to redistribute the information from higher dimension among the discrete ($\{s_i\}_{i=1}^N$), continuous ($\{x_i\}_{i=1}^N, H$) degrees of freedom. This form of learning is often referred to as unsupervised learning or classification.

As complexity of the embedded data increases, it is unfeasible to design metrics from scratch. We believe that the appropriate MDS design starts with the definition of adaptive and local (determined by s_i) geometry, such as that defined by Eq.(7). Our attempt was, from methodological viewpoint, partly inspired by theoretical framework of general relativity and geometrodynamics, where the fundamental postulate is made that geometry is determined by the mass-energy distribution analogous to distribution and structure of high dimensional dataset as counterpart.

In the field of MDS research, we would like to mention at least two approaches to "distance metric learning" or distance adaption particularly close in motivation to our approach. In [37] the distance is replaced by the linear function with the parameters determined by the regression. In the second approach [38] the monotonic nondecreasing function of the distance has been introduced in order to make the differences of distances less significant. In addition, our approach can be considered as being in line with the class of the adaptive ML approaches discussed in [39]. The results from implementation of *metric learning* approaches should be mentioned as well [40].

II. COMBINATION OF PARTICULAR OPTIMIZATION STRATEGIES

Below we present the heuristic optimization algorithm tailored to solve the MDS problem with adaptive metrics. The algorithm stops at time t^* and consists of the following subsequent steps (enumerated by t)

Step 1: GS optimization in polar coordinates

To refine the optimum estimation locally, we use polar grid mesh around randomly localized $\mathbf{x}_{i_{\text{rand}}}$ with i_{rand} drawn uniformly randomly from the set $\{1, 2, \dots, N\}$. The mesh is created with the radial step resolution Δr and angular step resolution $2\pi/N_n$. The mesh parameters Δr are drawn uniformly randomly from the respective interval $\langle \Delta r_D, \Delta r_U \rangle$. Then in the special case considered here $P = 2$ the algorithm generates the mesh of $N_n^2 N_s$ nearest-neighbor polar grid points

$$\begin{aligned} x_{i_{\text{rand}}, 1}^{\text{cand}}(l_r, l_\phi) &= x_{i_{\text{rand}}, 1} + \frac{\Delta r l_r}{N_n} \cos\left(\frac{2\pi l_\phi}{N_n}\right), \\ x_{i_{\text{rand}}, 2}^{\text{cand}}(l_r, l_\phi) &= x_{i_{\text{rand}}, 2} + \frac{\Delta r l_r}{N_n} \sin\left(\frac{2\pi l_\phi}{N_n}\right), \\ s_{i_{\text{rand}}}^{\text{cand}}(l_s) &= l_s. \end{aligned} \tag{8}$$

Within the standard logic of GS approach, the projections $\{x_{i_{\text{rand}}, z}^{\text{cand}}(l_r, l_\phi), z = 1, 2\}$, denoted by the superscript 'cand', represent candidate solutions of the respective optimization problem. Then, the candidate projections are

enumerated by the triplets

$$\{(l_r, l_\phi, l_s); l_r, l_\phi = 1, 2, \dots, N_n; l_s = 0, 1, \dots, N_s - 1\}. \quad (9)$$

In the case of feasibly high N_s and N_n , one can explore all the possible categories of $s_{i_{\text{rand}}}$ combinatorially. Obviously, the calculation of $V_{i_{\text{ran}}}^{\text{cand}} = V_{i_{\text{ran}}}$ using Eq.(4) must be preceded by reevaluation of the distances from $\mathbf{x}_{i_{\text{ran}}}^{\text{cand}}, s_{i_{\text{rand}}}^{\text{cand}}$ to all the other points. Let the coordinates $x_{i_{\text{rand}},1}^{\text{cand}}(l_{r,\text{min}}, l_{\phi,\text{min}}), x_{i_{\text{rand}},2}^{\text{cand}}(l_{r,\text{min}}, l_{\phi,\text{min}}), s_{i_{\text{rand}}}^{\text{cand}}(l_{s,\text{min}})$ correspond to the lowest local value of $V_{i_{\text{rand}}}^{\text{cand}}$. As this value is calculated using Eq.(4), its calculation must include changes in $e_{i_{\text{rand}},j}$ and $d_{i_{\text{rand}},j}^{(\text{P})}$ which are needed to update the $\mathbf{x}_{i_{\text{rand}}}$ and $s_{i_{\text{rand}}}$ values used in further optimization iterations.

Step 2: EO in the space of categorical variables

The optimization step is accepted with the decaying probability $\exp(-t/t_{\text{dec}})$ suggested to decay in time with the characteristic time constant t_{dec} . The strategy is similar to simulated annealing approach. At each algorithmic step, the instant worst part $i_{\text{max}} \in \{1, 2, \dots, N\}$ of the system defined by the respective maximum $V_{i_{\text{max}}} = \max_{i \in \{1, 2, \dots, N\}} V_i$ is localized. Then, the categorical variable $s_{i_{\text{max}}}$ is replaced by the value of s_i drawn randomly from the set $\{0, 1, \dots, N_s - 1\}$.

Step 3: HO - hysteresis along the variable H

Let's denote the best estimate of the optimum of the stress that algorithm attained by the time $(t-1)$ as $V_{\text{tot,best}}(t-1)$ and the total instant tension (potential) attained in time t as $V_{\text{tot}}(t)$, both calculated using Eq. (5) at the respective times. Let $V_{\text{tot,best}}(t-1)$ corresponds to $H_{\text{best}}(t-1)$ estimate of $H(t^*)$. Then, if $V_{\text{tot,best}}(t-1) > V_{\text{tot}}(t)$, the algorithm updates the $V_{\text{tot,best}}(t)$ and $H_{\text{best}}(t)$ as

$$H_{\text{best}}(t) \leftarrow H(t), \quad V_{\text{tot,best}}(t) \leftarrow V_{\text{tot}}(t). \quad (10)$$

Otherwise, previously obtained values are used to update the $V_{\text{tot,best}}$ and H_{best} , respectively

$$H_{\text{best}}(t) \leftarrow H_{\text{best}}(t-1), \quad V_{\text{tot,best}}(t) \leftarrow V_{\text{tot,best}}(t-1). \quad (11)$$

The HO dynamics is driven by the periodic exogenous signal

$$H_{\text{per}}(t) = H_D + \frac{H_U - H_D}{2} \left[1 + \cos\left(\frac{2\pi t}{t_{\text{per}}}\right) \right] \quad (12)$$

with the properly chosen period t_{per} . The signal $H_{\text{per}}(t)$ represents oscillations bounded by the H_D, H_U constants. The algorithm applies the strategy of the sequential linear mixing of the best estimate of the optimum with the decaying oscillations. The mixing is characterized by the coefficient $\exp(-t/t_{\text{dec}})$. The mixing process is incorporated into the non-autonomous recurrent dynamic rule in the form

$$H(t+1) = \underbrace{H_{\text{per}}(t) + (H_{\text{best}}(t) - H_{\text{per}}(t)) \left[1 - \exp\left(-\frac{t}{t_{\text{dec}}}\right) \right]}_{\text{tends to } H_{\text{best}}(t) \text{ as } t \rightarrow \infty}. \quad (13)$$

We remark, that an analogous learning strategy has been followed earlier [41] to solve the problem of Monte Carlo localization of the critical point under the noisy conditions. By other words, the formula is designed to reduce coupling between $H(t+1)$ and $H_{\text{per}}(t)$ under increasing influence of $H_{\text{best}}(t)$. It is straightforward to expect, that the ability to localize (hopefully global) optimum $H(t^*+1) \sim H_{\text{best}}(t^*)$ needs approximate fulfilment of the condition $t^* \gg t_{\text{dec}} \gg t_{\text{per}}$.

III. NUMERICAL EXPERIMENT

To illustrate here proposed algorithm, we apply it to several datasets drawn from analytically specified low dimensional manifolds embedded in an ambient space.

Firstly, we specify the list of parameters, which are common for all the system optimizations. The MDS with the projection onto $P = 2$ dimensions has been performed for $t^* \sim 3.0 \times 10^6$ iteration steps, but smaller number of iterations is sufficient for achieving comparable quality of the 2D projections. In the most cases, we consider systems including three data categories $N_s = 3$, thus $s_i \in \Omega = \{0, 1, 2\}$ (in the specific cases we used $N_s = 4, 5, 6$; the detailed specification is given in the corresponding figure captions). The optimization has been done for the search parameters $N_n = 10$ and $H_D = -1.5$, $H_U = 1.5$ (the bounds $H_D = -3.5$, $H_U = 3.5$ were also used to verify stability of obtained results). To perform the GS strategy, the mesh size was left to fluctuate within the bounds $\Delta r_D = 0.001$, $\Delta r_U = 0.3$. The dynamics of $H(t)$ was determined by the exogenous signal characterized by the parameters $t_{\text{dec}} = 3 \times 10^5$ and $t_{\text{per}} = 24 \times 10^3$. The optimization has been initialized from $x_{i,0} = X_{i,0}$, $x_{i,1} = X_{i,1}$ with the small additive noise, but the numerical experiments revealed that the initial conditions have only negligible influence on 2D projections. The tendencies of the heuristics and convergence towards the optimum has been controlled by monitoring of $V_{\text{tot}}(t)$. Any parametric approach requires parameter estimation. Despite many parameters to tune, we observed surprising robustness of the presented method in most cases as well as in different situations. The experience has shown us that what we need to focus on is the choice of parameters t_{dec} , t_{per} , H_D and H_U which seems to us play a key role in the determination of the minima.

The constructiveness of the algorithm is shown on the example of dataset. First, we considered system of $N = 300$ data items embedded into $D = 6$ space. The data were drawn from the parameterization

$$\begin{aligned} X_{i,1} &= \left[1 + \frac{1}{2} \cos \left(\frac{16\pi i}{N} \right) \right] \cos \left(\frac{2\pi i}{N} \right), \\ X_{i,2} &= \left[1 + \frac{1}{2} \cos \left(\frac{16\pi i}{N} \right) \right] \sin \left(\frac{2\pi i}{N} \right), \\ X_{i,3} &= \frac{1}{2} \sin \left(\frac{16\pi i}{N} \right), \\ X_{i,4} &= a_m \delta_{0,i \bmod 3}, \quad X_{i,5} = a_m \delta_{1,i \bmod 3}, \quad X_{i,6} = a_m \delta_{2,i \bmod 3}. \end{aligned} \tag{14}$$

The above dataset is constructed as a combination of the *toroidal spiral* (coordinates $X_{i,1}, X_{i,2}, X_{i,3}$) modified by geometric effects added by Kronecker delta and modulo functions $\delta_{j,i \bmod 3}$ calculated for the $j = 0, 1, 2$ components. The variable a_m is used to study different optimization conditions.

In Appendix I we present formulas for generating supplementary artificial datasets which outline consequences of the proposed method. In order to gain preliminary understanding of datasets, the relations between data pairs of Cartesian coordinates is plotted in Fig.1. In addition, to evaluate and facilitate the understanding of our method we compared results for PCA, classical metric MDS and diffusion map. The results are shown in Fig.2. Let us to note that in the case of PCA and MDS we used `princomp()` and `cmdscale()` R's base functions from stats package. For the implementation of diffusion map we used R function `diffuse()` from diffusionMap [42] package.

Let us focus on the problem of spiral studied for different a_m . The calculations (see the optimization results in Figs. 3 and 4 and the corresponding configurations in Fig. 5) revealed, that large a_m ($a_m \gg 1$) enhances the segregation process due to higher impact of the modular data structure and smaller influence of the harmonic functions forming

the 6D spiral. One of the most interesting findings is, that qualitative differences and regimes (see for example $a_m \sim 1$ cases presented in Fig.4 and Fig.5) may make the MDS analysis of some specific datasets more difficult than others.

The optimization of Klein bagel is presented in the Fig.6. Qualitatively, the optimization scenario is similar to the 6D modulated spiral, as well as to other simulations that we performed. The optimized projections of the half-sphere, Klein bagel, and full sphere samples are depicted in Figs. 7, 8 and 9, respectively. We see that these projections exhibit different levels of segmentation and compactness. Although the sample of Klein bagel resists partitioning into the compact regions of different categories, its spiral motif becomes more clear and plainly visible (see Fig.2) after the application of proposed method.

In addition to studies of artificial datasets we propose the explanation for empirical observations and their similarities. We focus on the epidemiological data of Hodgkin lymphomas for United states in the 2009-2010. The mortality data [43] contain absolute death counts by age splitted into five race dimensions: white (1), black (2), Asian/pacific islander (3), American Indian/Alaska native (4), Hispanic (5) ($D = 5$). The significance of the race of the patient for determination of the risk and efficiency of treatment has been discussed in [44, 45]. Before the application of MDS, the values on each particular dimension have been standardized to have zero mean and unit variance. Comparison of the application of classical MDS with our approach is depicted in Fig.10. Interestingly, both the 2D mappings show, that the observations can be embedded into one-dimensional manifold, which confirms the salient role of to age-related disease incidence. It also reflects the fact that time instants can be arranged in a one-dimensional manifold. In addition to classical MDS, our method also identified specificity of the categories belonging to the age bands 20-25, 30-40. It means, that the adaptive metrics enables to detect even small decline in the disease occurrence.

All the above examples lead to questions regarding the role of the number of categories, N_s . Thus, it would be interesting to mention manifolds which evidently cannot be mapped onto the plane. In other words, the intrinsic dimension is too high. As a simple example demonstrating this property may serve the maps of vertices of 6-dimensional hypercube, which are projectable onto 2D only on the expense of very high stress values (when the categorization absents). On the other hand, when the categorization via s_i is applied to the sample of \mathbb{R}^6 , the layered or slice projection structures are generated (see Fig.11). In agreement with intuitive expectations, the minima of V_{tot} deepen with the increase of N_s .

The following conclusions can be drawn from the numerical examples:

- (i) the system dynamics exhibits qualitative universal features which are independent from the investigated datasets;
- (ii) the "collective" coordinates $H(t), V_{\text{tot}}(t)$ resemble the phase portraits of the forced double-well harmonic oscillators subject to strong noisy disturbances due to EO presence. Interestingly, the occurrence of two local minima (one deep and one more shallow) seems to be a generic feature common to wide class of data. The exception is found in the uniform case $N_s = 1$, where the dependence on H and s_i simply vanishes.
- (iii) during the initial search phase $V_{\text{tot}}(t)$ suddenly drops. The subsequent adjustment yields gradual refinement of the double-minima structure.
- (iv) the local, slow convergence with slow detailed search is typical for the last optimization stage (in agreement with the optimization model, its assumptions and expectations, see the dynamics Eq.(13)).
- (v) in the most of investigated cases deeper minimum corresponds to negative $H(t^*)$. It means, that our algorithm

tends to interpret inter-category distances as smaller than $d_{i,j}^{(P=2)}$ (due to $1 + H|s_i - s_j|$ factor). This can be explained by the requirement to find sufficiently big area for the projections of the most of dense data inputs.

Interesting question arises whether oscillations of the parameters, such as those which determine $H_{\text{per}}(t)$, can be replaced by the complex dynamical models. The promising candidate for the alternative HO optimization part is the chaotic discrete Duffing oscillator [46] which was used in our numerical experiments. Our choice to use chaos has been motivated by the works which show increased optimization efficiency of the numerical sequences generated by means of the chaotic maps when compared to the random sequences [47]. As an example we used the Duffing map $x(t+1) = y(t)$, $y(t+1) = -0.2x(t) + 2.75y(t) - [y(t)]^3$, where $x(t+1)$ has been used to substitute the role of $H_{\text{per}}(t)$ ($y(t)$ is the auxiliary variable and the constants $-0.2, 2.75$ were chosen to belong to the chaotic regime). Chaos is often present in the nonlinear systems, thus many another variants of chaotic dynamical systems can be used to improve our MDS approach. Our preliminary simulations did not confirm increased efficiency in comparison to the presented harmonic stimulation. To assess the relevance of the chaotic models for the complex optimization ML problem one needs further intensive numerical research that goes far beyond the scope of the present work. In any case, the application of the idea of self-organization in combination with chaos phenomena may constitute very interesting way to develop further ML studies.

IV. DISCUSSION

ML plays an important and growing role in exploratory data analysis and machine learning. In the paper we introduced biologically and physically inspired flexible variant of standard MDS, which, as we have shown, is the effective tool for simultaneous mapping and categorization. Our proposal uses the orchestration of three stochastic optimization heuristics. We demonstrated, that optimization trajectory produces the phase portrait involving two well-separated minima. This monitoring properly illustrates and, at the same time, justifies the necessity of the more comprehensive and advanced routines in the optimization.

Current ML theories are handled in more or less linear framework with too small influence of non-linearity to exploit emergent characteristics. The main contribution of the present paper (in comparison with traditional ML theories) consists in combining reasonable solution of the specific problems with promising non traditional approximate heuristic methods used in the area of the complex systems, statistical physics, optimization science and artificial biology. Although not practical for the applications in which the projections and partitions must be found rapidly, the approach seems to be successful at more detailed analysis of the selected manifolds. We have shown, that the flexibility of MDS can be improved using the adaptive metrics containing categorical independent variables. Overall, surprising results clearly reveal unexpected behavior and delineate the domains where this robust technique can bring valuable results. Increased flexibility has to be paid with a larger number of categories. Although the current form of the algorithm enables categorization of the datasets into predefined number of categories, further research is needed to tune the number to match the intrinsic dimension and local structure of manifolds. The information criteria (such as BIC) have to be included as well to improve the overall classification performance.

Non-equilibrium dynamical systems often show complex adaptive behavior called emergent properties. In the most of the datasets that we studied the emergence of the compact domains of the categories has been observed. It should be emphasized that processes, where structural changes of initial disordered configurations/projections yield self-

organized structures, significantly differ from the traditional forms of the programming and learning which claim to produce similar effects by using some explicit and user predefined criteria.

A lot of open questions remains to be studied in the proposed computational scheme. A general open question is whether we should generalize the method to include information about more generic distance functions and more nuanced interpretations and categorizations. Our analysis does not rule out other modified forms of the $1 + H|s_i - s_j|$ prefactor of the metrics [see Eq.(7)]. The model we used is flexible enough to admit straightforward extensions. For example, the prefactor $1 + H|s_i - s_j| + H_{\text{symmetry viol.}}(s_i + s_j)$ may be used which violate the original symmetry of $|s_i - s_j|$ (i.e. symmetry between an object s_i and its mirror $N_s - 1 - s_i$) by adding, e.g., the term $H_{\text{symmetry viol.}}(s_i + s_j)$ that prevent from the occurrence of non-uniqueness and degeneracy.

Our analysis leads to interesting application of the hysteretic optimization method which has relevance in modeling and adjusting of systems with global impact parameters. We have shown that the examples can be found in ML situations where high intrinsic dimension of manifold favors the charting of data during the projection process. In particular, we believe that our numerical experimentation might be instructive in the construction of the models of the collective behavior including features such as emergence and organization, which are topics of much interest in the current research. We foresee further potential applications to variety of ML problems that are formulated as optimization tasks.

The authors would like to gratefully acknowledge Project CELIM (316310) "Fostering Excellence in Multiscale Cell Imaging" funded by European Community Seventh Framework Program FP7 EU, and European X-Ray Laser Project XFEL.

V. APPENDIX - LIST OF THE SYNTHETIC DATA STRUCTURES

The appendix describes the list of three parametrizations I, II, III of 3d manifolds, used to generate datasets suitable for MDS variant of ML numerical experiments.

I. Half - Sphere ($D = 3$ approximated by $N = 17^2 = 289$ data items)

The data are generated using

$$\begin{aligned} X_{i,1} &= \cos \phi_j \cos \theta_k, & X_{i,2} &= \sin \phi_j \cos \theta_k, \\ X_{i,3} &= \sin \theta_k, \\ \theta_k &= \pi(k/17), & \phi_j &= 2\pi(j/17). \end{aligned} \tag{15}$$

Here θ_k and ϕ_j denotes the sequences (samples) in azimuthal and polar coordinates, respectively. The index $i \equiv i(j, k)$ represents the enumeration mark of (j, k) elements of the Cartesian product

$$CP_I \equiv \{ (j, k); \ j = 0, 1, \dots, 16; \ k = 0, 1, \dots, 16 \}. \tag{16}$$

Thus for example: $i(0, 0) = 1$ (it means that here $j = 0, k = 0$), $i(0, 1) = 2$ (here $j = 0, k = 1$), $i(0, 2) = 3, \dots$ $i(1, 0) = 18, i(1, 1) = 19, i(1, 2) = 20, \dots i(16, 14) = 287, i(16, 15) = 288, i(16, 16) = 289$. The analogous notation is used in the case of datasets II and III. Note that sample of the full sphere which is projected in Fig.(9) is created

by replacements $\cos \theta_k \rightarrow \sin \theta_k$, $\sin \theta_k \rightarrow \cos \theta_k$. It means that data object we call "full sphere" includes the same number of data inputs as the data object "half sphere".

II. Möebius strip ($D = 3$; $N = 16 \times 17 = 272$ data items; we examined also denser data variant $N = 462 = 22 \times 21$);

$$X_{i,1} = [1 + (v_j/2) \sin(u_k/2)] \cos u_k, \quad (17)$$

$$X_{i,2} = [1 + (v_j/2) \cos(u_k/2)] \sin u_k,$$

$$X_{i,3} = (v_j/2) \sin(u_k/2),$$

$$v_j = 1 - 2(j/17), \quad u_k = 2(k/17)\pi. \quad (18)$$

Again $i \equiv i(j, k)$ enumerates $CP_{II} \subset CP_I$, where $CP_{II} \equiv \{ (j, k); j = 0, 1, \dots, 16; k = 1, 2, \dots, 16 \}$ slightly differs from CP_I .

III. Klein bagel ($D = 3$; $N = 289$ data items; denser dataset variant includes $N = 484$ items;)

The manifold is homeomorphic to the well known Klein bottle. The data are generated using

$$X_{i,1} = [a_R + \cos(\theta_j/2) \sin v_k - \sin(\theta_j/2) \sin(2v_k)] \cos \theta_j, \quad (19)$$

$$X_{i,2} = [a_R + \cos(\theta_j/2) \sin v_k - \sin(\theta_j/2) \sin(2v_k)] \sin \theta_j,$$

$$X_{i,3} = \sin(\theta_j/2) \sin v_k + \cos(\theta_j/2) \sin(2v_k),$$

$$\theta_j = 2\pi j/17, \quad v_k = 2\pi k/17. \quad (20)$$

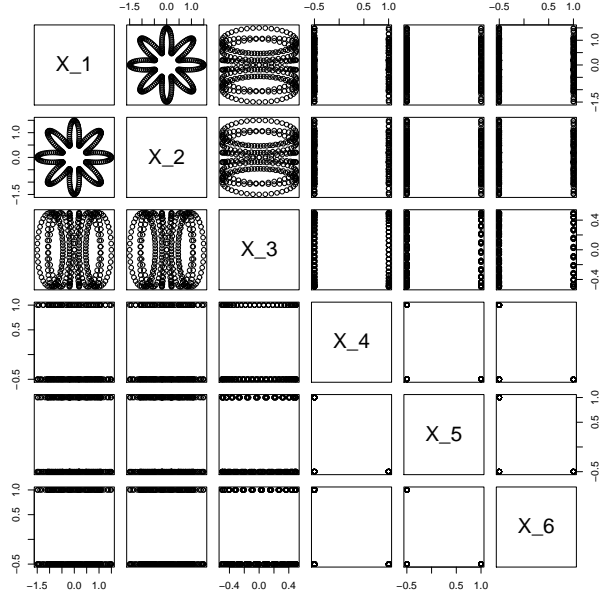
The model depends on the single parameter we choose $a_R = 1.5$. Here $i \equiv i(j, k)$ enumerates set $(j, k) \in CP_I$.

-
- [1] Y. Ma and Y. Fu. *Manifold Learning Theory and Applications*. CRC Press, 2011.
 - [2] H. Suetani, K. Soejima, R. Matsuoka, U. Parlitz, and H. Hiroki. Manifold learning approach for chaos in the dripping faucet. *Phys. Rev. E*, 86:036209, Sep 2012.
 - [3] P. Schwander, D. Giannakis, Ch. H. Yoon, and A. Ourmazd. The symmetries of image formation by scattering. ii. applications. *Optics Express*, 20(12):12827–12849, 2012.
 - [4] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1989.
 - [5] N. T. Trendafilov and I. T. Jolliffe. Projected gradient approach to the numerical solution of the scotlass. *Computational Statistics & Data Analysis*, 50(1):242253, January 2006.
 - [6] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
 - [7] J. B. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2):115–129, 1964.
 - [8] T. Cox and M. Cox. *Multidimensional Scaling*. Chapman & Hall, London, 1994.
 - [9] T.F. Cox. Multidimensional scaling in process control. *Handbook of Statistics*, 22:609–623, 2003.
 - [10] J. D. Carroll, P. Arabie, and L.J. Hubert. *K.Kempf-Leonard (Ed.), Encyclopedia of Social Measurement*, chapter Multidimensional Scaling (MDS), pages 779–784. Elsevier, San Diego, 2005.
 - [11] L. Chen and A. Buja. Stress functions for nonlinear dimension reduction, proximity analysis, and graph drawing. *Journal of Machine Learning Research*, 14:1145–1173, 2013.

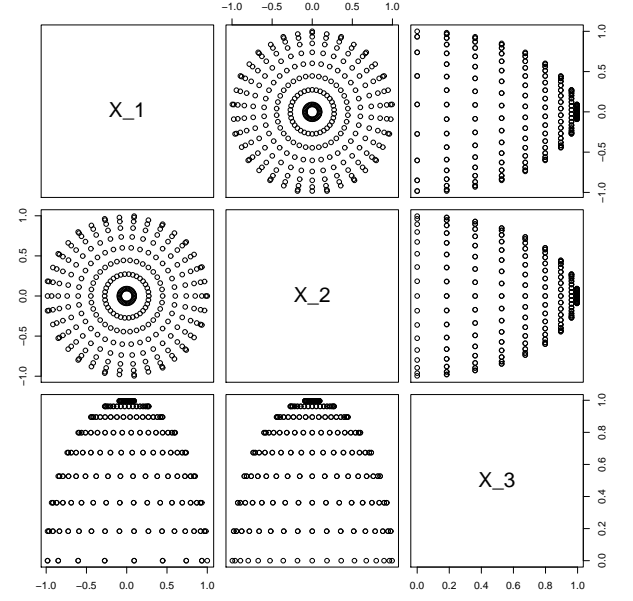
- [12] J. B. Tenenbaum, Vin de Silva, and J. C. Langford. Global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.
- [13] H. Zha and Z. Zhang. Continuum isomap for manifold learnings. *Computational Statistics & Data Analysis*, 52(1):184–200, September 2007.
- [14] J. Yina, D. Hua, and Z. Zhoua. Noisy manifold learning using neighborhood smoothing embedding. *Pattern Recognition Letters*, 29(11):16131620, 2008.
- [15] A. K. H. Duc, M. Modat, K. K. Leung, M. J. Cardoso, J. Barnes, T. Kadir, and S. Ourselin. Using manifold learning for atlas selection in multi-atlas segmentation. *PLoS ONE*, 8(8):e70059, 2013. Available at <http://dx.doi.org/10.1371/journal.pone.0070059>.
- [16] J.W. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18:401–409, 1969.
- [17] O. Kramer and F. Gieseke. Evolutionary kernel density regression. *Expert Systems with Applications*, 39:9246–9254, 2012.
- [18] A. N. Gorban and A. Zinoviev. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, chapter Chapter 2: Principal Graphs and Manifolds Algorithms, Methods and Techniques, pages 28–59. IGI Global, ISR, August 2009.
- [19] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis. Diffusion maps, spectral clustering and eigenfunctions of fokker-planck operators. *Neural Information Processing Systems (NIPS)*, 18, 2005.
- [20] B. Nadler, S. Lafon, R.R. Coifman, and I.G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, 2006.
- [21] J. X. Li. Visualization of high-dimensional data with relational perspective map. *Information Visualization*, 3(1):49–59, 2004.
- [22] R. Xiao, Q.Zhao, D.Zhang, and P. Shi. Facial expression recognition on multiple manifolds. *Pattern Recognition*, 44:107–116, 2011.
- [23] A. Noack. Energy models for graph clustering. *Journal of Graph Algorithms and Applications*, 11(2):453–480, 2007.
- [24] A. Noack. Modularity clustering is force-directed layout. *Phys. Rev. E*, 79, 2009.
- [25] B. Alatas. Chaotic bee colony algorithms for global numerical optimization. *Expert systems with Applications*, 37:5682–5687, 2010.
- [26] S. Binitha and S.S. Sathya. A survey of bio inspired optimization algorithms. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(2):2231–2307, 2012.
- [27] A. Biswas, K. K. Mishra, S. Tiwari, and A. K. Misra. Physics-inspired optimization algorithms: A survey. *Journal of Optimization*, 2013, 2013.
- [28] R. Horst and P.M. Pardalos, editors. *Handbook of Global Optimization*. Kluwer Academic Publishers, Dordrecht, 1995.
- [29] S. Boettcher. Extremal optimization of graph partitioning at the percolation threshold. *J. Phys. A: Math. Gen.*, 32(28):5201–5211, 1999.
- [30] G. Zarand, F. Pazmandi, K.F.Pal, and G.T. Zimanyi. Using hysteresis for optimization. *Phys. Rev. Lett.*, 89(15):150201, 2002.
- [31] P. Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Computing in Science and Engineering*, 71(24):4083–4086, 1993.
- [32] S. Boettcher. Extremal optimization: heuristics via coevolutionary avalanches. *Computing in Science and Engineering*, 2(6):75–82, 2000.
- [33] T. Zhou, W.J. Bai, L. J. Cheng, and B.H. Wang. Continuous extremal optimization for lennard-jones clusters. *Phys. Rev. E*, 72(1):016702, 2005.
- [34] S. Boettcher, A.G. Percus, and M. Grigni. Optimizing through co-evolutionary avalanches. In J.-P. Finance, editor, *Parallel Problem Solving from Nature PPSN VI*, number 1917 in Lecture Notes in Computer Science, pages 447–456.

Springer-Verlag, 2000.

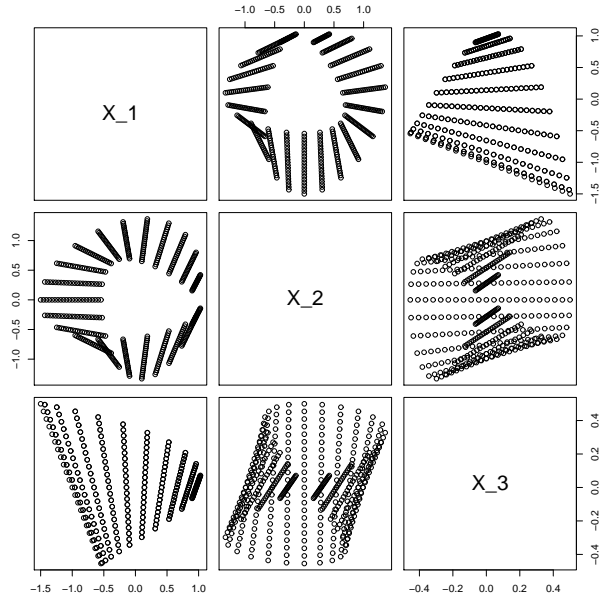
- [35] K. F. Pal. Hysteretic optimization for the Sherrington - Kirkpatrick spin glass. *Physica A: Statistical Mechanics and its Applications*, 367:261–268, 2006.
- [36] K. F. Pál. Hysteretic optimization for the traveling salesman problem. *Physica A: Statistical Mechanics and its Applications*, 329(1-2):287–297, 2003.
- [37] L. Kari, K. A. Hill, A.S. Sayem, N. Bryans, K.Davis, and N.S. Dattani. Map of life: Measuring and visualizing species-relatedness with "molecular distance maps". 2013.
- [38] E.A. Cansizoglu, M. Akcakaya, U. Orhan, and D. Erdogmus. Manifold learning by preserving distance orders. *Pattern Recognition Letters*, 38:120–131, 2014.
- [39] H. Yin. Advances in adaptive nonlinear manifolds and dimensionality reduction. *Front. Electr. Electron. Eng. China*, 1(6):72–85, 2011.
- [40] L. Zhang, L. Zhang, D. Tao, X.Huang, and B. Du. Hyperspectral remote sensing image subpixel target detection based on supervised metric learning. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8):4955–4965, 2014.
- [41] D. Horvath and M.Gmitra. The self-organized multi-lattice monte carlo simulation. *Int. J. Mod. Phys. C*, 15(09):1249–1268, 2004.
- [42] Joseph Richards. *diffusionMap: Diffusion map*, 2014. R package version 1.1-0.
- [43] Centers for Disease Control and Prevention (CDC), 1999-2010. Available at <http://apps.nccdc.cdc.gov/USCS>, Public-use data file.
- [44] A. Zaki, N. Natarajan, and C.J. Mettlin. Early and late survival in hodgkin disease among whites and blacks living in the united states. *Cancer*, 72(2):602–606, Jul 1993. <http://www.ncbi.nlm.nih.gov/pubmed/8319194?dopt=Abstract>.
- [45] Lymphoma Research Foundation. Hodgkin Lymphoma (HL), cited May 2014. Available at <http://www.lymphoma.org/site/pp.asp?c=bkLTKa0QLmK8E&b=6300137>.
- [46] O. Junge, J. E. Marsden, and I. Mezic. Uncertainty in the dynamics of conservative maps. In *Decision and Control, CDC. 43rd IEEE Conference*, volume 2, pages 2225–2230. IEEE, 2004.
- [47] B. Li and W.S. Jiang. Chaos optimization method and its application. *Journal of Control theory and Applications*, 14(4):613–615, 1997.



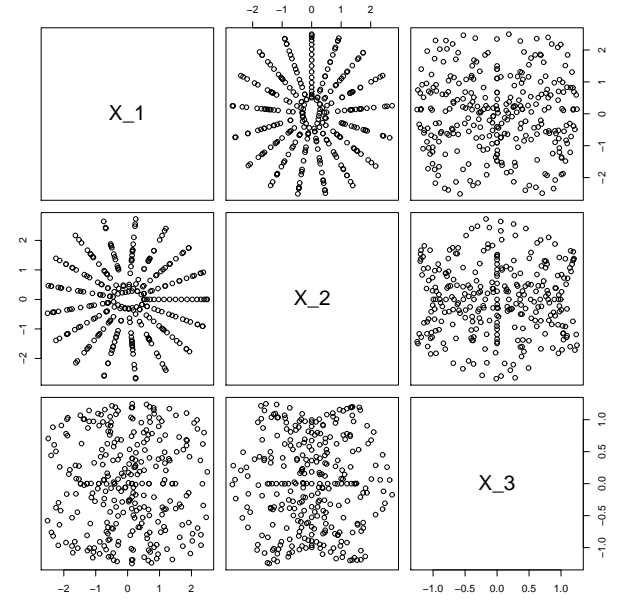
(a) spiral



(b) half sphere



(c) Möeb. st.



(d) Klein b.

FIG. 1: The multiple scatterplot attempt to represent the relationships among variables of four selected artificial datasets. The data for 6D modulated toroidal spiral are drawn from Eq.(14) for the parameter $a_m = 1.2$.

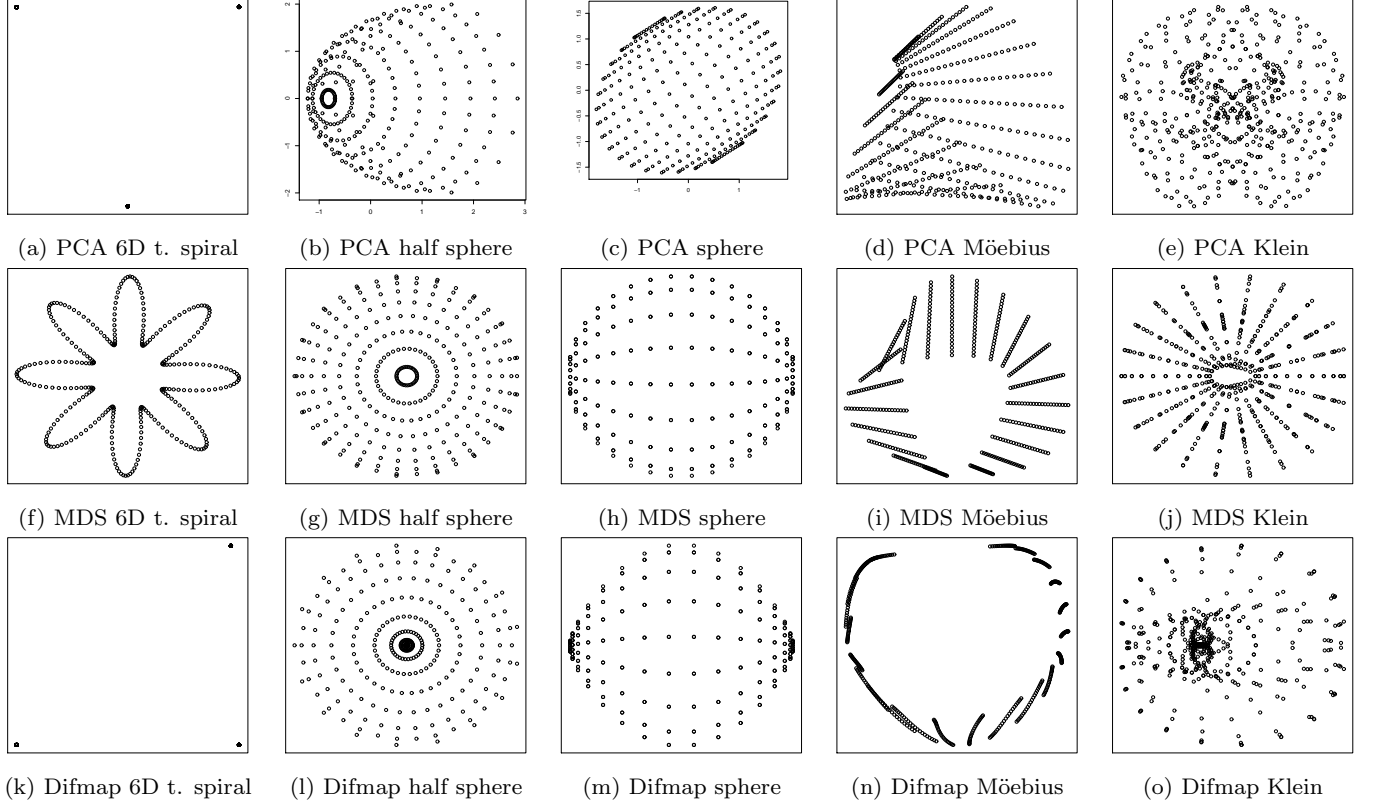


FIG. 2: The comparison of the PCA, MDS and diffusion map is done for several generated artificial datasets. It should be noted that diffusion maps rely on the choice of the tuning parameter ϵ that dictates the threshold for similarity. We present the results for $\epsilon = 0.1$. Surprisingly, many of the phase portrait patterns are reminiscent of those found in Fig.1. The most remarkable differences between outputs of the mentioned methods have been detected for 6D toroidal spiral ($a_m = 1.2$) and Klein bagel.

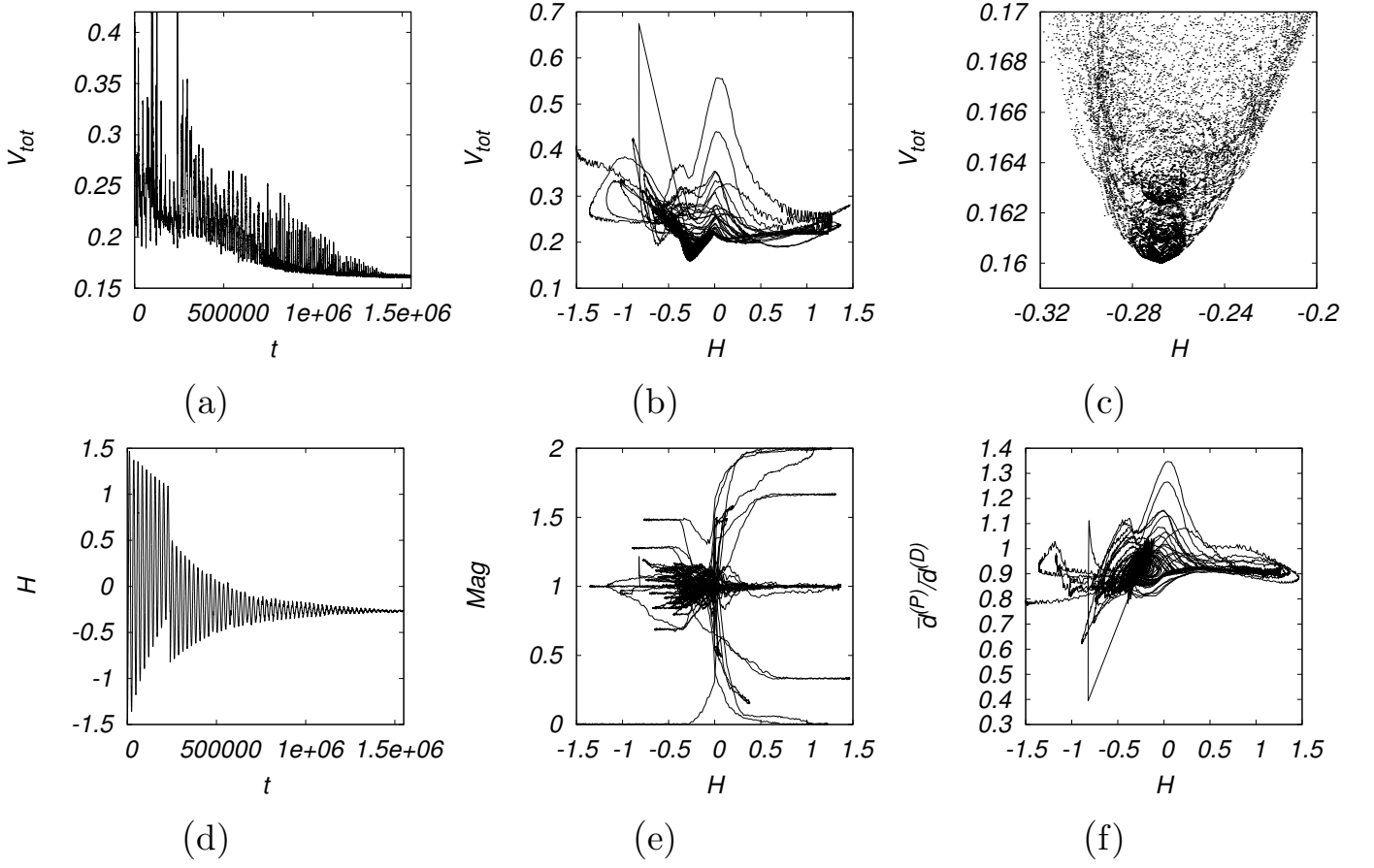


FIG. 3: The MDS optimization represented by phase portraits in terms of collective variables. The formation of MDS projection of the original data sampled from 6D toroidal spiral defined by Eq.(14); the artificial data calculated for $a_m = 1.5$. The dependences show remarkable effect of the exogenous oscillations. The figure parts inform how the particular variables vary during optimization process. The part (a) demonstrates global decrease of $V_{tot}(t)$. The visible are random excitations (avalanches) which stem from the application of EO. The figure parts (b), (c) depict the parametric plot of the total stress $V_{tot}(t)$. They admit to estimate the optimal $H(t^*)$ that corresponds to $V_{tot}(t^*) \sim -0.27$. The part (d) depicts the illustrative break in the decaying exponential regime of $H(t)$ caused by the self-organized HO. The part (e) shows that at the minimum the instant $Mag(t)$ saturates around 1; (f) The alternative confirmation of the convergence of the optimization: the optimized ratio of the mean distances of the original and projected coordinates tends to the value slightly less than 1.

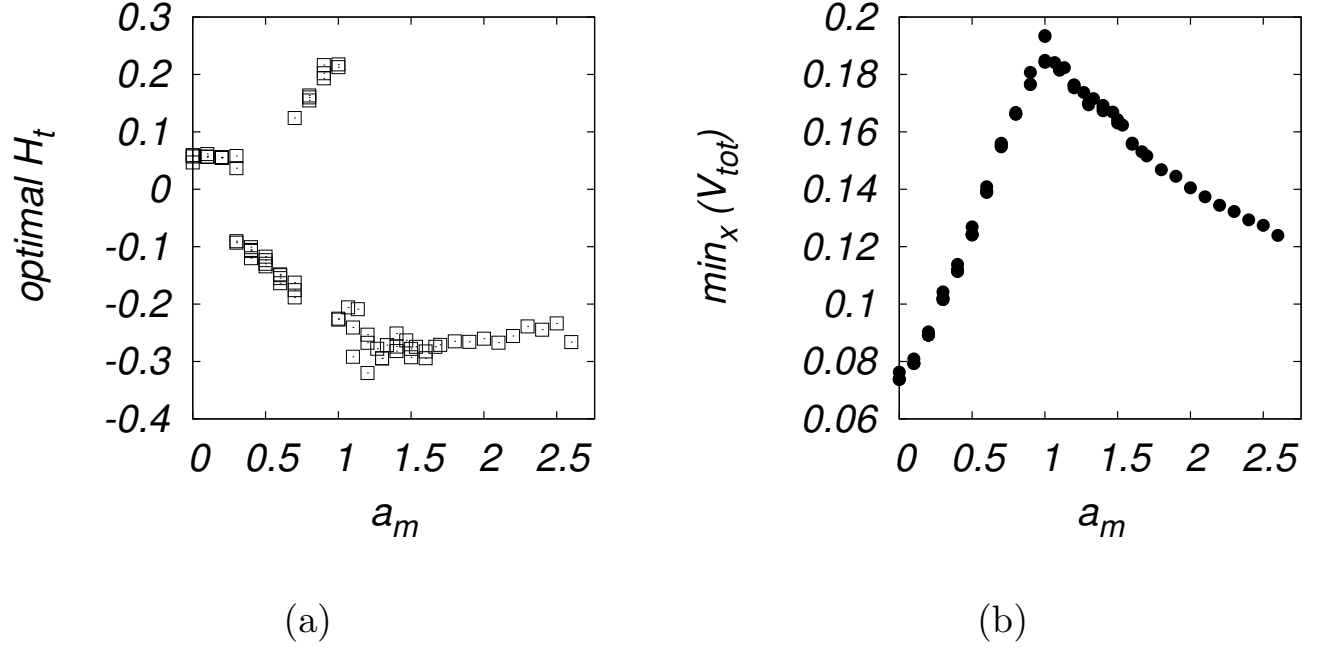


FIG. 4: The systematics of the of optimal characteristics of the projections of 6D toroidal spiral. The results obtained for different a_m (see Eq.(14)). We see that data may also exhibit critical-like properties of nonlinear changes. The part (a) shows several breaks in the a_m dependence of the optimal $H(t^*)$. Two branches of the competing solutions differ in the sign of $H(t^*)$. The part (b) shows qualitative change indicated by the cusp-peak of the optimal $V_{tot}(t^*)$ localized near to $a_m \sim 1$. The peak appears to be related to the structural change, which is viewable in the corresponding configuration ($a_m \simeq 1$) in Fig.5.

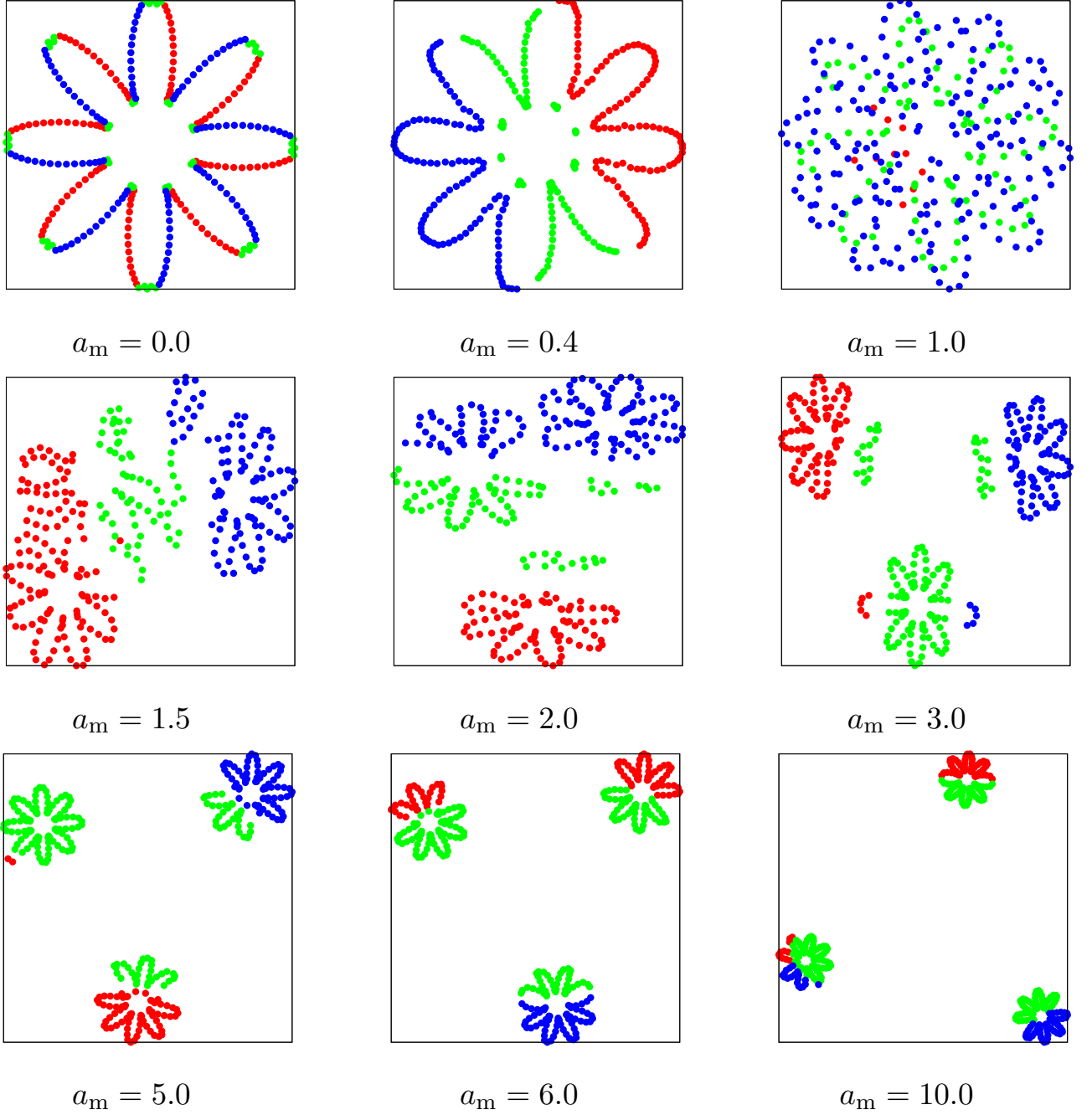


FIG. 5: The nearly optimal configurations of the spiral projections obtained for different a_m values. The ternary like structure forms due to presence of the modulo function. Moreover, it is rather logical to observe that separation between three visible parts increases with a_m . The results suggest different degree of the regularity and different quality of results as well for different a_m .

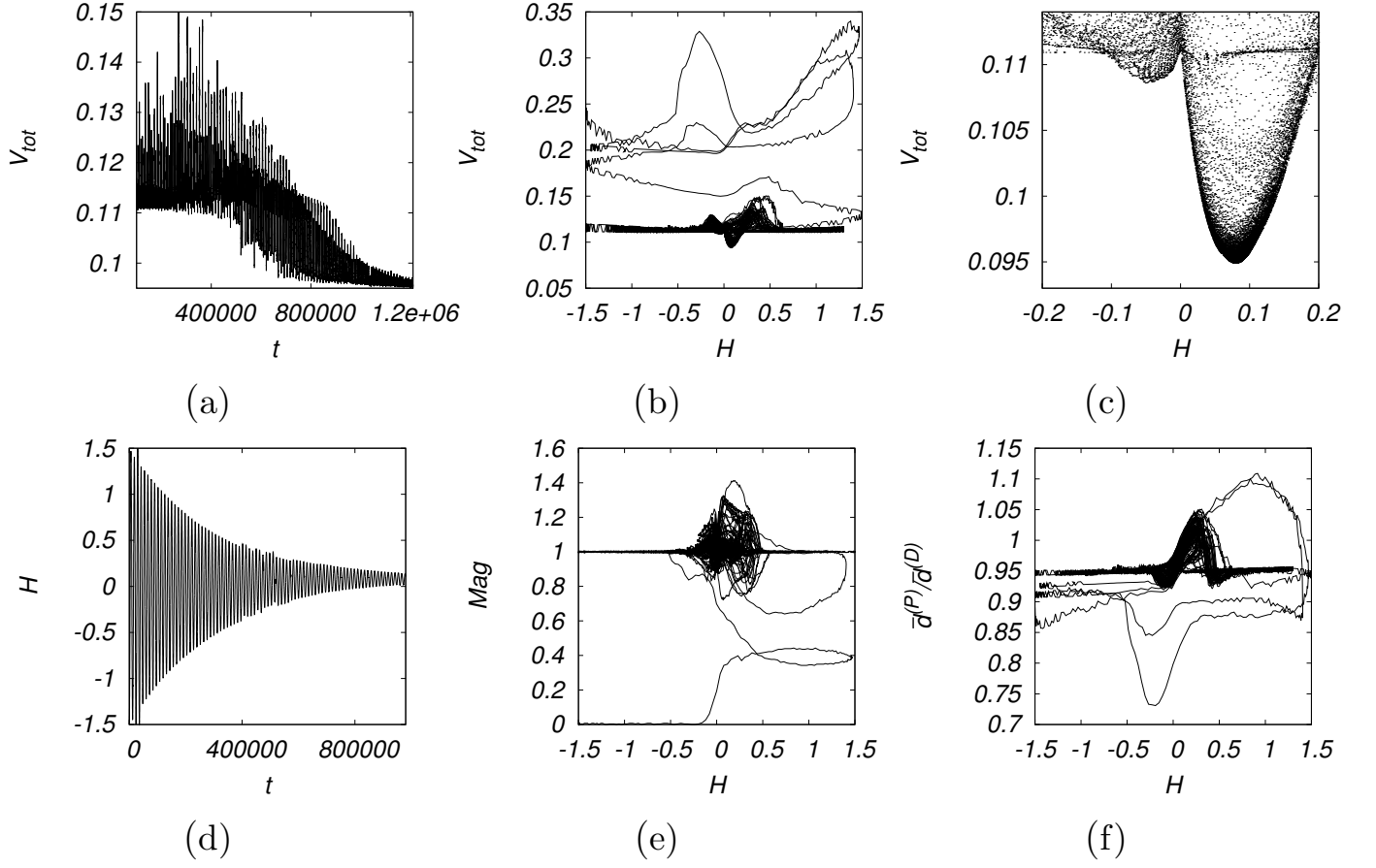


FIG. 6: The monitoring of the complex optimization dynamics, which yields 2d projection of the Klein bagel. The results highlight the inherent complexity of the optimization dynamics. The stochastic attempts of overcoming of local barriers are clearly visible. See Fig.3 for more detailed comments of the partial phase portraits.

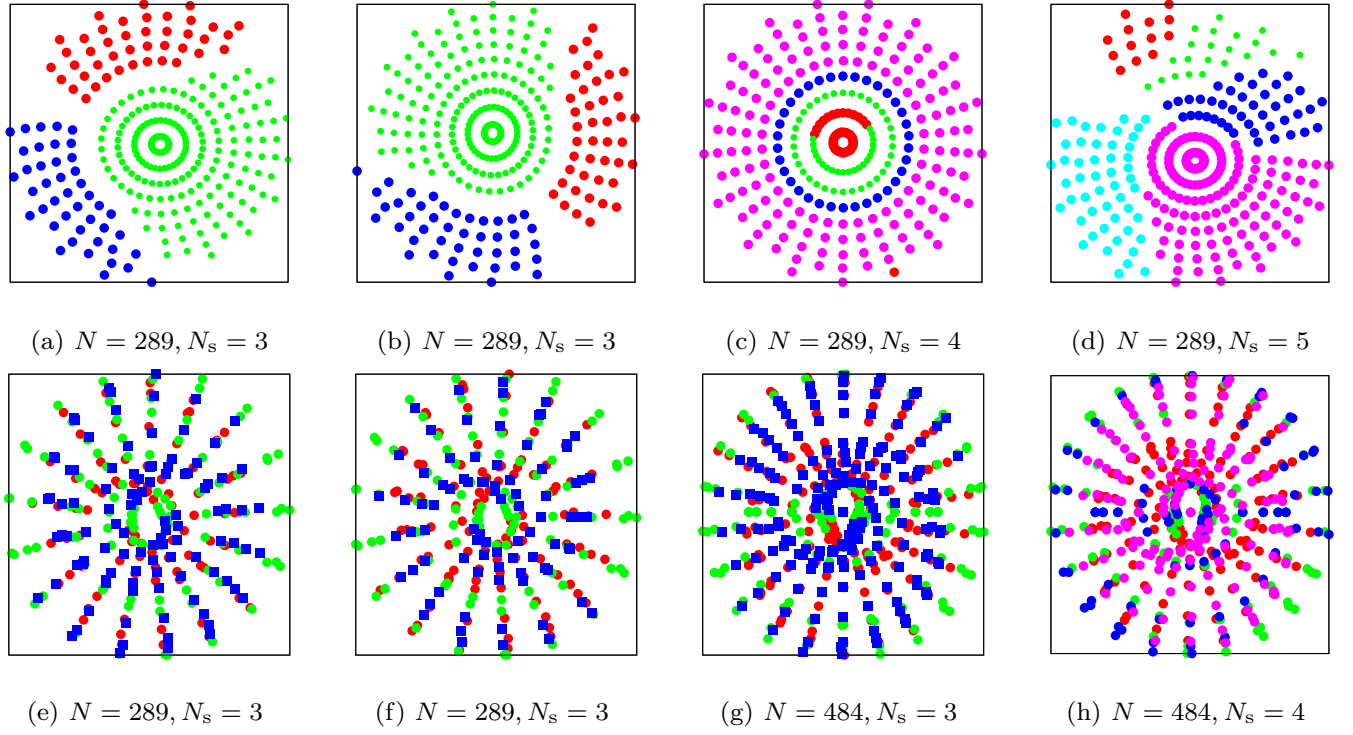


FIG. 7: The variability or multi-fold degeneracy (in the physical sense) of the 2D projections. The alternative charts (each belonging to the specific unique categorical variable s_i) of 3D half-sphere: see the parts (a), (b), (c), (d). The results obtained for the Klein bagel are plotted in the parts (e), (f), (g), (h). The projections are calculated for different initial conditions and stochastic realizations. The identical colors correspond to the categories: *red* ($s_i = 0$), *green* ($s_i = 1$), *blue* ($s_i = 2$) or *pink* ($s_i = 4$). The mapping uncovered some key differences between the comprehensibility of the sphere and Klein bagel. The mapping and formation of the charts in the case of sphere is relatively well understandable. On the other hand, the interlinked and nested spiral structures corresponding to different categories represent the only projectable information about the Klein bagel.

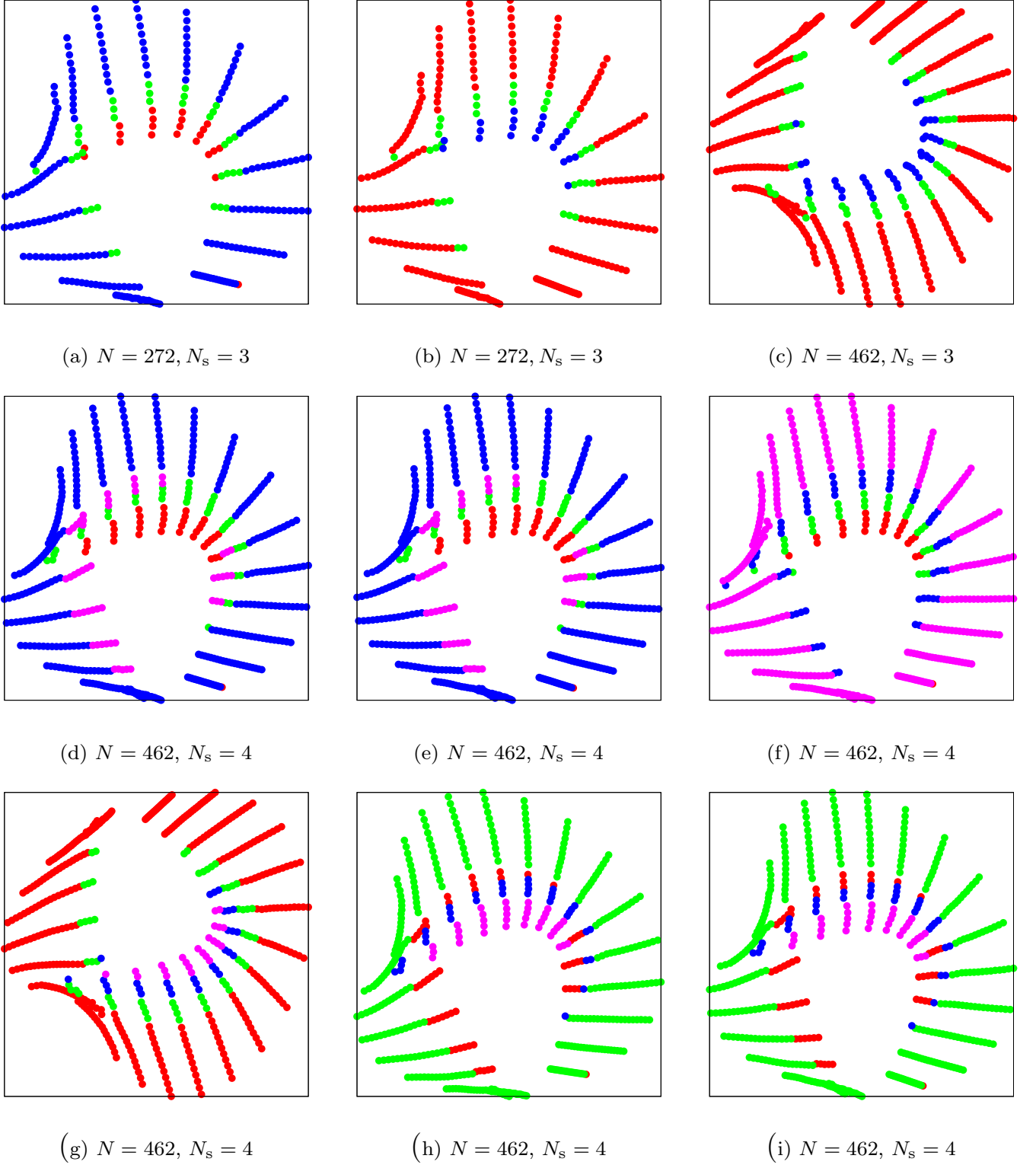


FIG. 8: The alternative projections of the Möbius strip calculated for different initial conditions. Part (a): see the emergence of the middle strip which improves the orientation in the visualization of the structure. The denser data (c) better describe the higher variability, but they do not reveal some excessive differences. Note that difference between $s_i = 0$ and $s_j = 2$ domains (accompanied by the color difference) is only apparent since the locations of the classes 0, 2 are replaceable because of their unique metric inter-class distance multiplicative factor $\sim 1 + H|s_i - s_j| = 1 + 2H$ (see Eq.(7)).

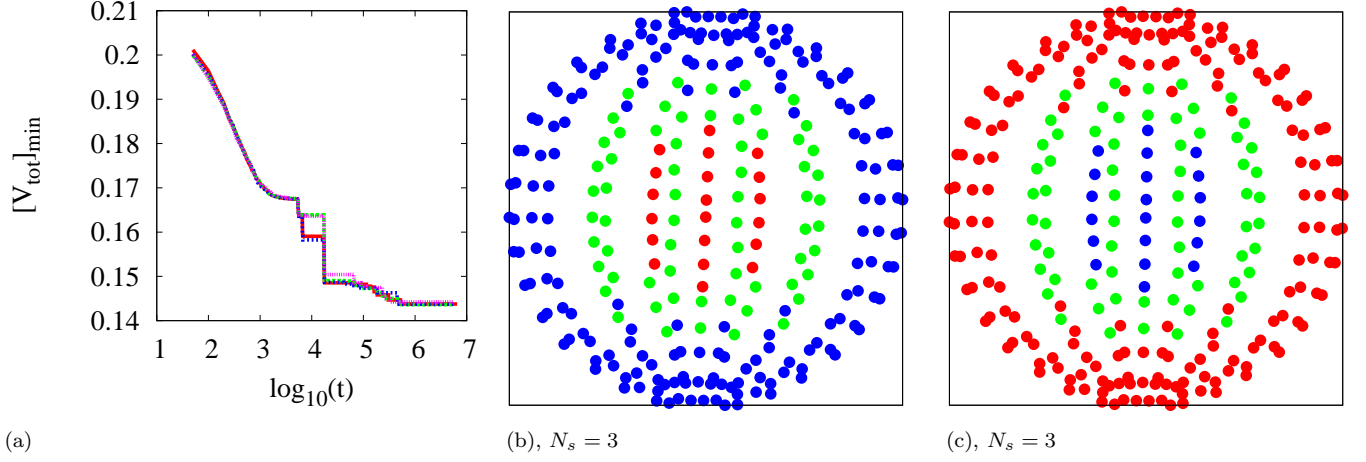


FIG. 9: The demonstration of the two-fold degeneracy in the case of the categorization of the data drawn from the sphere. Our version of MDS is applied to data drawn from the 3D sphere. The part (a) represent the actual minimum is recalculated for four independent simulation stochastic runs. Two examples - parts (b) and (c) of the configurations that exhibit the invariance with respect to spin transformations: $s = 0$ (red) $\rightarrow s = 2$ (blue) ; $s = 1$ (green) $\rightarrow s = 1$ (green). The emergence of distinguishing characteristics (see alternating vertical lines) of the front and back surface of sphere.

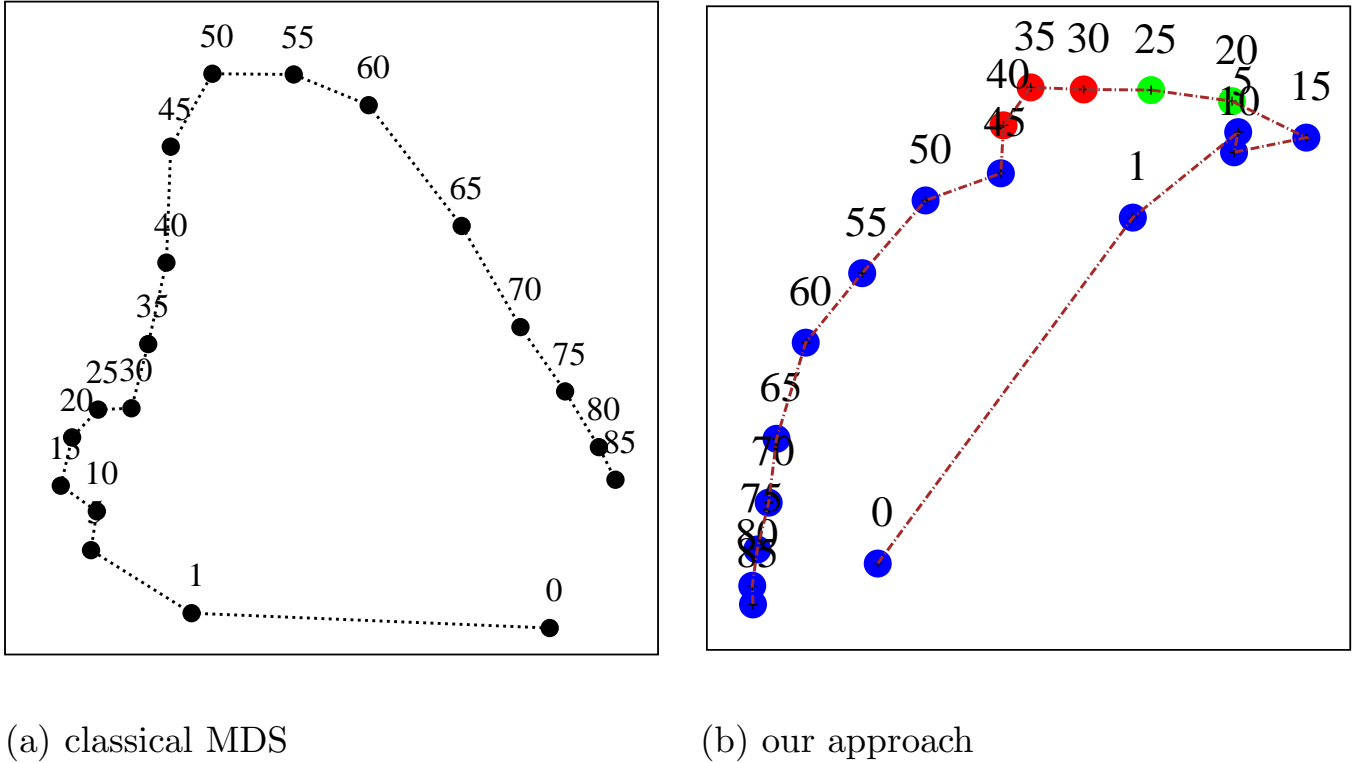


FIG. 10: The MDS results obtained for 5-dimensional race-specific data for Hodgkin's lymphoma data. The comparison of classical metric MDS and our approach (which provides optimum $H = -0.064$, $V_{\text{tot}} = 0.026$). The effect of young - old age similarity closing the loop is remarkable. The key advantage of our methods is that it identifies the significant changes in the categories of the middle-aged adults.

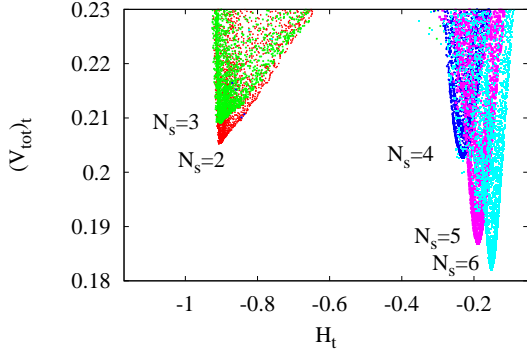
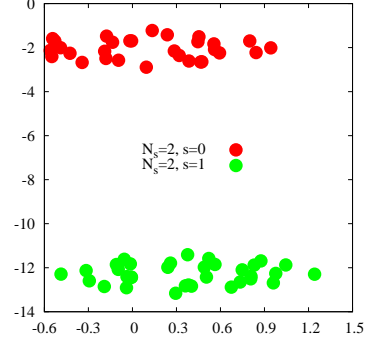
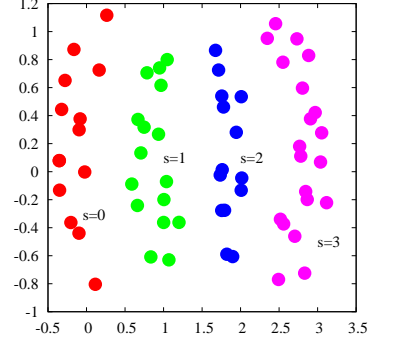
(a) $N_s = 6$ (b) $N_s = 2$ (c) $N_s = 4$

FIG. 11: The influence of N_s integer parameter on the depth of the minimum V_{tot} achieved (see part a) for dataset constructed from the vertices of 6-dimensional cube. The emergence of layered structures presented in the parts (b) and (c).